methods among which poisson and negative binomial regression are the most popular (Washington et al., 2011). In a poisson regression model, the probability of site i having y_i accidents per year (where y_i is a non-negative integer) is given as follows:

$$P(y_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!} \qquad (1)$$

where $\lambda_i = f(\beta X_i)$ (2)

X_i is the set of site characteristics, β represents the set of coefficients that need to be estimated, and *f* is a function that relates the site characteristics to the poisson parameter λ_i (expected number of crashes per year at the site). The limitation of the poisson distribution is that the mean and variance are considered equal. Most often with crash data, the variance has been found to exceed the mean. This phenomenon is called overdispersion. Negative binomial regression is able to account for this overdispersion by allowing the variance to differ from the mean as follows:

$$Var(y_i) = E(y_i) + k[E(y_i)]^2 \qquad (3)$$

where k is the overdispersion parameter, Var is the variance and E is the expected value (i.e., mean). Negative binomial regression has become the most common method for developing SPFs and is also the recommended modeling approach in the HSM.

In both poisson and negative binomial regression, the most common function *f* is the exponential function. In other words, the relationship can be written as follows:

$$\lambda_i = \exp(\beta X_i) \qquad (4)$$

The exponential function implies a log-linear relationship between site characteristics and the expected number of crashes per year at the site (i.e., $\ln(\lambda_i) = \beta X_i$, where ln represents the natural logarithm). The log-linear relationship has become common because it allows the poisson and negative binomial regression models to be estimated using a technique called generalized linear models (McCullagh and Nelder, 1989). In fact, all the predictive models reported in Part C of the HSM assumed a log-linear relationship between site characteristics and the expected number of crashes. More recently, some researchers have argued that other functional forms (other than the log-linear relationship) need to be investigated since the log-